On the uniqueness of probability matching priors

By Ana-Maria Staicu

Department of Statistics, University of Toronto, Toronto, Ontario, Canada, M5S 3G3

anamaria@utstat.toronto.edu

and Nancy Reid

Department of Statistics, University of Toronto, Toronto, Ontario, Canada, M5S 3G3

reid@utstat.toronto.edu

SUMMARY

Probability matching priors are priors for which Bayesian and frequentist inference, in the form of posterior quantiles, or confidence intervals, agree to some order of approximation. These priors are constructed by solving a first order partial differential equation, that may be difficult to solve. However, Peers (1965) and Tibshirani (1989) showed that under parameter orthogonality a family of matching priors can be obtained. The present work shows that, when used in a third order approximation to the posterior marginal density, the Peers-Tibshirani class of matching priors is essentially unique.

Some key words: approximate Bayesian inference; Laplace approximation; orthogonal parameters; tail probability approximation.

1 Introduction

We consider parametric models for a response $Y = (Y_1, \ldots, Y_n)^T$ with joint density $f(y;\theta)$. The parameter $\theta^T = (\psi, \lambda^T)$ is assumed to be a d-dimensional vector with ψ the scalar parameter of interest. The log-likelihood function is $\ell(\theta) = \ell(\theta; y) = \log f(y;\theta)$. Let $j(\theta) = -n^{-1}\ell_{\theta^T\theta}(\theta; y)$ be the observed information matrix and $i(\theta) = n^{-1}E\{-\ell_{\theta^T\theta}(\theta; Y); \theta\}$ the expected Fisher information matrix per observation.

In the absence of subjective prior information about the parameter θ, it may be natural to use a prior which leads to posterior probability limits that are also frequentist limits in the sense that:

$$\mathrm{pr}_\pi\{\psi \leq \psi^{(1-\alpha)}(Y) \mid Y\} = \mathrm{pr}_\theta\{\psi^{(1-\alpha)}(Y) \geq \psi\} + O(n^{-1})$$

where $\psi^{(1-\alpha)}(Y)$ is the upper $(1-\alpha)$ quantile of the marginal posterior density $\pi_m(\psi \mid Y)$. Following Datta & Mukerjee (2004) we call such priors first order probability matching priors. In a model with a scalar parameter, Welch & Peers (1963) showed that $\pi(\theta) \propto i^{1/2}(\theta)$ is the unique first order probability matching prior.

In models with nuisance parameters, we partition the observed and expected information matrices in accordance to the partition of the parameter; for the expected Fisher information matrix we have the elements

$$i_{\psi\psi}(\theta), \qquad i_{\psi\lambda}^T(\theta) = i_{\lambda\psi}(\theta), \qquad i_{\lambda\lambda}(\theta).$$

Peers (1965) derived a class of first order matching priors for ψ, as solutions to a partial differential equation. See also Mukerjee & Ghosh (1997), who provided a simpler derivation. In general this differential equation is not easy to solve, unless the components ψ and λ are orthogonal with respect to expected Fisher information, i.e. $i_{\psi\lambda}(\theta) = 0$. In this case Tibshirani (1989) and Nicolau (1993) show that a family of solutions is:

$$\pi(\psi, \lambda) \propto i_{\psi\psi}^{1/2}(\psi, \lambda) g(\lambda), \tag{1}$$

where $g(\lambda)$ is an arbitrary function. Sometimes consideration of higher order matching enables restriction of the class of functions $g(\lambda)$, occasionally enabling a unique matching prior to be defined; see Mukerjee & Dey (1993).

Levine & Casella (2003) propose solving the partial differential equation numerically, in models with a single nuisance parameter. Sweeting (2005) considers vector nuisance parameters and introduces data-dependent priors that locally approximate the matching priors; these are used in conjunction with a Metropolis-Hastings procedure. Both methods require substantial computational effort. Our work is closely connected to DiCiccio & Martin (1993), who use matching priors in approximate Bayesian inference as an alternative to more complicated frequentist formulas.

In the present paper we show that when a Laplace type approximation to the posterior distribution is used, all priors of the form (1) lead to the same posterior inference. If an orthogonal parameterization is not explicitly available, the differential equations defining parameter orthogonality can be used in conjunction with (1) to give an expression for the prior in the original parameterization. We use the invariance argument presented in Mukerjee & Ghosh (1997) to express the matching prior in terms of the original parameterization. In §2 we present the Laplace approximation to the marginal posterior, and in §3 give the results on uniqueness. In §4 we discuss models where the orthogonal components can be obtained without solving the differential equations, arguing that the proposed approach reduces the degree of difficulty. In §5 we illustrate the results through several examples taken from Datta & Ghosh (1995), Sweeting (2005) and Levine & Casella (2003).

2 Approximate Bayesian inference

If $\pi(\psi, \lambda)$ is the prior for θ, Bayesian inference for ψ is based on the marginal posterior density $\pi_m(\psi \mid y)$, and the Laplace approximation to this is given by:

$$\pi_m(\psi \mid y) \doteq c|j_p(\hat{\psi})|^{1/2} \exp\{\ell_p(\psi) - \ell_p(\hat{\psi})\} \left\{ \frac{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|}{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|} \right\}^{1/2} \frac{\pi(\psi, \hat{\lambda}_\psi)}{\pi(\hat{\psi}, \hat{\lambda})},$$

where $\hat{\lambda}_\psi$ is the constrained maximum likelihood estimate, $\ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi)$ is the profile log-likelihood for ψ, $\hat{\theta}^T = (\hat{\psi}, \hat{\lambda}^T)$ is the full maximum likelihood estimate, and $j_p(\hat{\psi}) = -\ell_p''(\hat{\psi})$ is the observed information corresponding to the profile log-likelihood. In the independently and identically distributed sampling context Tierney & Kadane (1986) showed that the Laplace approximation has relative error $O(n^{-3/2})$.

The corresponding $O(n^{-3/2})$ approximation to the marginal posterior tail probability is:

$$\mathrm{pr}_\pi(\Psi \geq \psi \mid Y) = 1 - \Pi_m(\psi \mid y) \doteq \Phi(r) + \left(\frac{1}{r} - \frac{1}{q_B}\right)\phi(r) \tag{2}$$

where ϕ and Φ are standard normal density and standard normal distribution function respectively, and

$$\begin{aligned} r &= \mathrm{sign}(\hat{\psi} - \psi)[2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\}]^{1/2} \\ q_B &= \ell_p'(\psi)\{j_p(\hat{\psi})\}^{-1/2} \left\{ \frac{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|}{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|} \right\}^{1/2} \frac{\pi(\hat{\psi}, \hat{\lambda})}{\pi(\psi, \hat{\lambda}_\psi)} \end{aligned} \tag{3}$$

(DiCiccio & Martin, 1991).

3 First order probability matching priors

3.1 Orthogonal parameterization

When the model is given in an orthogonal parameterization the first order matching prior for the parameter of interest ψ is given by (1). It enters approximation (2) as a ratio so the relevant quantity is:

$$\frac{i_{\psi\psi}^{1/2}(\hat{\psi}, \hat{\lambda})g(\hat{\lambda})}{i_{\psi\psi}^{1/2}(\psi, \hat{\lambda}_\psi)g(\hat{\lambda}_\psi)}.$$

Although the function $g(\lambda)$ is an arbitrary factor in (1), for sufficiently smooth g the ratio $g(\hat{\lambda})/g(\hat{\lambda}_\psi) = 1 + O_p(n^{-1})$ as a consequence of the result that $\hat{\lambda}_\psi = \hat{\lambda} + O_p(n^{-1})$ under parameter orthogonality. It follows that the approximation to $\Pi_m(\psi \mid y)$ in (2) is unique to $O(n^{-1})$. Having a unique approximation to the marginal posterior probabilities to $O(n^{-1})$ leads to unique posterior quantiles for ψ to $O_p(n^{-3/2})$. This can be verified by the technique of inversion of asymptotic series. The details are outlined in the Appendix.

Another approximation to the marginal posterior, also accurate to error $O(n^{-3/2})$, is the Barndorff-Nielsen approximation discussed in DiCiccio & Martin (1991):

$$1 - \Pi_m(\psi \mid y) \doteq \Phi\{\ r - r^{-1}\ \log(r/q_B)\ \}; \tag{4}$$

this version gives an explicit expression for the quantiles, which are also invariant to the choice of $g(\cdot)$ in (1) to this order.

We call the prior

$$\pi_U(\psi, \lambda) \propto i_{\psi\psi}^{1/2}(\psi, \lambda) \tag{5}$$

the "unique matching prior" for the component ψ, under the orthogonal parameterization ψ and λ. This uniqueness was noted in DiCiccio & Martin (1993), although with a different approach and interpretation, but seems to have been overlooked in Casella & Levine (2003) and Sweeting (2005).

3.2 General parameterization

Assume our model is given in a parameterization $\phi^T = (\psi, \eta^T)$ not necessarily orthogonal, and denote by $\theta^T = (\psi, \lambda^T)$ an orthogonal reparameterization. Since ψ is scalar component, such a parameterization always exists, and is given as a solution to the partial differential equation:

$$i_{\psi\eta}(\phi) = \frac{\partial \lambda^T(\phi)}{\partial \psi} \left\{ \frac{\partial \lambda^T(\phi)}{\partial \eta} \right\}^{-1} i_{\eta\eta}(\phi), \tag{6}$$

(Cox & Reid, 1987). The unique matching prior $\pi_U(\psi, \lambda)$ can be written in the original parameterization as:

$$\pi_U(\psi, \eta) \propto i^{1/2}_{\psi\psi.\eta}(\psi, \eta)\ J(\psi, \eta), \tag{7}$$

where $i_{\psi\psi.\eta}(\psi, \eta) = i_{\psi\psi}(\psi, \eta) - i_{\psi\eta}(\psi, \eta) \{i_{\eta\eta}(\psi, \eta)\}^{-1} i_{\eta\psi}(\psi, \eta)$ is the (ψ, ψ) component of the expected Fisher information in the orthogonal parameterization, and $J(\psi, \eta) = |\partial\lambda/\partial\eta^T|_+$ is the Jacobian of the transformation. In accordance with calling prior (5) a unique matching prior in the orthogonal parameterization (ψ, λ^T), the prior (7) shall be referred to as the unique matching prior in the (ψ, η^T) parameterization.

The analogy between (5) and (7) can be also justified by noting that in the orthogonal parameterization $\theta^T = (\psi, \lambda^T)$, the unique matching prior for ψ is proportional to the square root of the inverse of the asymptotic variance for $\hat{\psi}$. For a general parameterization $\phi^T = (\psi, \eta^T)$ the variance of $\hat{\psi}$ is the inverse of the partial information for ψ, i.e. $i^{\psi\psi}(\phi) = \{i_{\psi\psi.\eta}(\phi)\}^{-1}$ (Severini, 2000, Ch 3.6), so the matching prior (7) in parameterization ϕ is a natural extension of the unique matching prior (5). Note that although the orthogonal reparameterization of the original model parameterization is not unique, all solutions lead to the same expression for q_B to $O_p(n^{-1})$ and thus to the same posterior quantiles limits to $O_p(n^{-3/2})$.

The unique matching prior (7) is similar to the local probability matching prior proposed by Sweeting (2005). The two priors share the term involving the partial information $i_{\psi\psi.\eta}$; the extra factor in Sweeting's local prior is proportional to a local approximation of the Jacobian $J(\psi, \eta)$, based only on the parameter of interest and on the overall maximum likelihood estimate.

While $i_{\psi\psi.\eta}(\psi, \eta)$ is usually referred to as the partial information for ψ, we note that this applies strictly only for an orthogonal reparameterization. For an arbitrary interest-respecting transformation of θ, the (ψ, ψ) component of the expected Fisher information, expressed in the original parameterization, is

$$i_{\psi\psi}(\phi) - \frac{\partial \lambda^T(\phi)}{\partial \psi} \left\{ \frac{\partial \lambda^T(\phi)}{\partial \eta} \right\}^{-1} i_{\eta\eta}(\phi) \left\{ \frac{\partial \lambda(\phi)}{\partial \eta^T} \right\}^{-1} \frac{\partial \lambda(\phi)}{\partial \psi} \; ; \tag{8}$$

this simplifies to the expression for the partial information for ψ, only in the case of interest-respecting orthogonal reparameterization. This can be immediately derived from the positive definite nature of the information block matrix component $i_{\eta\eta}(\phi)$. If w^T denotes the row vector given by the right hand side of expression (6), then expression (8) equals $i_{\psi\psi.\eta}(\psi, \eta)$ if and only if any of the following equalities hold:

$$\begin{aligned} i_{\psi\eta}(\psi, \eta) \{i_{\eta\eta}(\psi, \eta)\}^{-1} i_{\eta\psi}(\psi, \eta) &= w^T \; \{i_{\eta\eta}(\psi, \eta)\}^{-1} \; w \\ \left\{ i^T_{\psi\eta}(\psi, \eta) - w \right\}^T \; \{i_{\eta\eta}(\psi, \eta)\}^{-1} \; \left\{ i^T_{\psi\eta}(\psi, \eta) - w \right\} &= 0 \end{aligned}$$

which reduces to $w = i^T_{\psi\eta}(\psi, \eta)$, since $\{i_{\eta\eta}(\psi, \eta)\}^{-1}$ is a positive definite matrix.

Another version of the Laplace approximation to the marginal posterior density for ψ can be obtained using the adjusted profile log-likelihood function, $\ell_a(\psi) = \ell_p(\psi) - \frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|$:

$$\pi_m(\psi \mid y) \doteq c|j_a(\hat{\psi})|^{1/2} \exp\{\ell_a(\psi) - \ell_a(\hat{\psi})\} \frac{\pi(\psi, \hat{\lambda}_\psi)}{\pi(\hat{\psi}, \hat{\lambda})},$$

where $j_a(\psi) = -\ell_a''(\psi)$. This approximation also has a relative error of $O(n^{-3/2})$, and to the same order can be integrated to give the approximate posterior distribution

$$1 - \Pi_m(\psi \mid y) = \int_\psi^\infty \pi_m(\psi \mid y) \doteq \Phi(r_a) + \left(\frac{1}{r_a} - \frac{1}{q_{Ba}}\right)\phi(r_a), \tag{9}$$

where

$$\begin{aligned} r_a &= \operatorname{sign}(q_{Ba})[2\{\ell_a(\hat{\psi}) - \ell_a(\psi)\}]^{1/2} \\ q_{Ba} &= \ell_a'(\psi)\{j_a(\hat{\psi})\}^{-1/2} \, \frac{\pi(\hat{\psi}, \hat{\lambda})}{\pi(\psi, \hat{\lambda}_\psi)}. \end{aligned} \tag{10}$$

Expression (10) corresponds to expression (17) for $\bar{T}$ from DiCiccio & Martin (1993), using the results that $j_a(\hat{\psi}) = j_p(\hat{\psi})\{1 + O_p(n^{-1})\}$ and $\hat{\psi}_a - \hat{\psi} = O_p(n^{-1})$, where $\hat{\psi}_a$ is the point that maximizes $\ell_a(\psi)$; see Reid (2003).

When the model is given in an orthogonal parameterization $\theta^T = (\psi, \lambda^T)$, similar arguments to those presented in §3.1 confirm that the matching prior (5) gives unique approximate Bayesian inference to $O(n^{-1})$ using approximation (9).

If the original parameterization is $\phi^T = (\psi, \eta^T)$ and an orthogonal reparameterization $\theta^T = (\psi, \lambda^T)$ has the property $\hat{\lambda}_\psi = \hat{\lambda}$, then the Cox-Reid adjustment becomes parameterization invariant under interest-respecting orthogonal reparameterization. Hence the matching prior (7) gives unique approximate Bayesian inference to $O(n^{-3/2})$ in the class of first order matching priors.

Rahul Mukerjee (personal communication) has showed that when $\hat{\lambda}_\psi = \hat{\lambda}$, the unique first order matching prior is second order matching if and only if the model has the property that

$$\frac{\partial}{\partial \psi}\left[i_{\psi\psi}^{-3/2}(\theta)E\{\ell_\psi^3(\theta)\}\right] = 0;$$

this condition is the same as the second order matching condition for the scalar parameter model. He has also showed that the same result holds under the weaker condition $E_\theta(\hat{\lambda}_\psi) - E_\theta(\hat{\lambda}) = o(n^{-1})$.

4 A note on parameter orthogonality

An orthogonal reparameterization may in some models be constructed by exploiting the property that if $\hat{\lambda}_\psi = \hat{\lambda}$ holds for all ψ then the components ψ and λ are orthogonal with respect to expected Fisher information (Barndorff-Nielsen & Cox, 1994, Ch 3.6). The property $\hat{\lambda}_\psi = \hat{\lambda}$ is referred to as "strong orthogonality".

First consider η to be a scalar nuisance parameter. If the score function corresponding to the nuisance parameter η has the form

$$\ell_\eta(\psi, \eta;\ y) \propto\ h\left\{\lambda(\psi, \eta);\ y\right\}, \tag{11}$$

for some functions $h(\cdot; y)$ and $\lambda(\cdot, \cdot)$ with $|\partial\lambda(\psi, \eta)/\partial\eta^T| \neq 0$, where the proportionality refers to non-zero functions which depend on the parameter only, then λ and ψ are strongly orthogonal. This follows from the equivariance of the constrained maximum likelihood estimator $\hat{\eta}_\psi$: $\ell_\eta(\psi, \hat{\eta}_\psi) = 0$ is equivalent to $h\left\{\lambda(\psi, \hat{\eta}_\psi);\ y\right\} = 0$, so any solution of this equation $\hat{\lambda}_\psi = \lambda(\psi, \hat{\eta}_\psi)$ depends on y only and not ψ, i.e. $\hat{\lambda}_\psi = \hat{\lambda}$.

This result can be extended to the case that the nuisance parameter is a vector;

the function h is then a vector of functions. Details are provided in the Appendix. We use strong orthogonality in Example 2 of §5.4. For models with strong orthogonality the difficulty of obtaining matching priors is reduced significantly.

A simple form of (11) frequently encountered is $h\{\lambda(\psi,\eta);\ y\} = \lambda(\psi,\eta) - \tilde{p}(y)$ where $|\partial\lambda(\psi,\eta)/\partial\eta^T| \neq 0$. Such is the case for the mean value reparameterization in the exponential family model. Another class of models giving strong orthogonality of parameters are those with likelihood orthogonality: i.e. $L(\psi,\eta) = L_1(\psi)\ L_2\{\lambda(\psi,\eta)\}$. The one-way random effects model in §5.3 belongs to this class.

5 Examples

5.1 Linear exponential family

Consider a sample of independently and identically distributed observations $Y = (Y_1,\ldots,Y_n)^T$ from the model:

$$f(y_i;\phi) = \exp\{\ \psi s(y_i) + \eta^T t(y_i) - c(\phi) - d(y_i)\} \tag{12}$$

where $\phi^T = (\psi,\eta^T)$ is the full parameter and ψ the component of interest. An orthogonal reparameterization is given by $\theta^T = (\psi,\lambda^T)$ with $\lambda = E_\theta\{t_+(y)\}$, where $t_+(y) = \sum_{i=1}^n t(y_i)$. This can be obtained from the orthogonality equation (6), but more directly by noting that the arguments of the previous section ensure that $\hat{\lambda}_\psi = \hat{\lambda}$. Approximation (2) with the prior (1) is independent of $g(\cdot)$ to $O(n^{-3/2})$, i.e. the matching prior $\pi_U(\psi,\lambda) \propto i_{\psi\psi}(\psi,\lambda)$ is unique; in the initial parameterization

$$\pi_U(\phi) \propto i^{1/2}_{\psi\psi.\eta}(\phi)\ |c_{\eta\eta}(\phi)|_+.$$

In approximation (2), the expression for q_B simplifies to:

$$q_B \;=\; \ell_\psi(\tilde{\phi})\; i_{\psi\psi.\eta}^{-1/2}(\tilde{\phi})\; \left\{ \frac{|i_{\eta\eta}(\hat{\phi})|}{|i_{\eta\eta}(\tilde{\phi})|} \right\}^{1/2},$$

where $i_{\psi\psi.\eta}(\phi) = c_{\psi\psi}(\phi) - c_{\psi\eta^T}(\phi)\left\{c_{\eta^T\eta}(\phi)\right\}^{-1} c_{\eta\psi}(\phi)$, $\tilde{\phi}^T = (\psi,\ \hat{\eta}_\psi^T)$ and $\hat{\phi}^T = (\hat{\psi},\ \hat{\eta}^T)$. The example is considered in DiCiccio & Martin (1993) as well.

5.2 Logistic regression

We analyze the urine data of Davison and Hinkley (1997, Example 7.8). The presence or absence of calcium oxalate crystals in urine as well as specific gravity, pH, osmolarity, conductivity, urea concentration and calcium concentration are measured for 77 complete cases. The relationship between calcium oxalate crystals and the 6 explanatory variables is investigated under the logistic regression model. Matching priors for logistic regression are obtained numerically in Levine & Casella (2003) and Sweeting (2005); here we give a simple analytical solution.

The logistic regression model for a vector of independent random variables $Y = (Y_1, \ldots, Y_n)^T$ is

$$\begin{aligned} Y_i|p_i &\sim \text{Binomial}(m_i, p_i), \qquad i = 1, \ldots, n \\ \text{logit}(p_i) &= \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi} \end{aligned}$$

where $\beta = (\beta_0, \ldots, \beta_p)^T \in \mathbb{R}^{p+1}$. The log-likelihood is:

$$\ell(\beta) = \sum_{i=1}^{n} y_i \left(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}\right) - \sum_{i=1}^{n} m_i \log\left\{1 + e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}}\right\}.$$

Assume the parameter of interest is $\psi = \beta_p$, and take $\eta = (\beta_0, \ldots, \beta_{p-1})^T$ be the nuisance parameter. Since the model is an exponential family, $\lambda = E_\beta\{t(y)\} = E_\beta(\sum_{i=1}^n y_i, \ldots, \sum_{i=1}^n y_i x_{p-1i})^T$ is orthogonal to ψ. Writing $V(\beta) = \text{diag}\{m_i p_i(1-p_i)\}$, we have

$$\begin{aligned} i_{\psi\psi}(\beta) &= x_p^T V(\beta) x_p \\ i_{\psi\eta}(\beta) &= x_p^T V(\beta) X_{-p} \\ i_{\eta\eta}(\beta) &= X_{-p}^T V(\beta) X_{-p}. \end{aligned}$$

where X is the $n \times (p+1)$ model matrix and $X_{-p} = X - \{x_p\}$ is the $n \times p$ matrix obtained by removing the column vector x_p.

Note also that $i_{\eta\eta}(\beta) = \partial\lambda/\partial\eta^T$. The unique matching prior has the form:

$$\pi_U(\beta) \propto i_{\psi\psi.\eta}^{1/2}(\beta)|i_{\eta\eta}(\beta)|_+, \tag{13}$$

with $i_{\psi\psi.\eta}(\beta) = i_{\psi\psi}(\beta) - i_{\psi\eta}(\beta)\{i_{\eta\eta}(\beta)\}^{-1} i_{\eta\psi}(\beta)$. When assessing inference for $\psi = \beta_p$, the Lugannani and Rice formula (2) is used with the entries r and q_B having the following expressions:

$$\begin{aligned} r &= \text{sign}(\hat\psi - \psi)[2\{\ell(\hat\beta)) - \ell(\tilde\beta)\}]^{1/2} \\ q_B &= \ell_\psi(\tilde\beta)\; i_{\psi\psi.\eta}^{-1/2}(\tilde\beta)\left\{\frac{|i_{\eta\eta}(\hat\beta)|}{|i_{\eta\eta}(\tilde\beta)|}\right\}^{1/2}, \end{aligned}$$

where for convenience we write $\tilde\beta$ for the constrained maximum likelihood estimate $(\hat\beta_{0,\psi}, \ldots, \hat\beta_{p-1,\psi})$ and $\hat\beta$ for the maximum likelihood estimate.

For illustration, we take $\psi = \beta_6$, the coefficient of the effect of calcium concentration on the presence of calcium oxalate crystals in urine. The 95% posterior probability intervals using the Bayesian approach with matching prior (13) are given in Table 1 and Figure 1. Also shown are two (first order) normal approximations and the third

	95% CI for β_6		p-value
Normal approximation to m.l.e. $\hat{\beta}_6$	(0.3169	1.250)	4.9887e-004
Normal approximation to conditional m.l.e. $\hat{\beta}_6^c$	(0.2631	1.160)	9.3724e-004
Third order frequentist approximation	(0.3224	1.208)	6.6893e-006
Laplace approximation with prior (13)	(0.3213	1.211)	5.3555e-006

Table 1: Comparison of 95% confidence intervals for β_6 and comparison of p-values for testing H_0: $\beta_6 = 0$

order approximation to the conditional distribution of $\sum_{i=1}^{n} x_{6i} y_i$ given $t = \hat{\lambda}$. The frequentist calculations were carried out using the `cond` package in the `hoa` library bundle for `R` (Brazzale, 2000). Although this package does not provide the Bayesian solution explicitly, the components needed are readily derived from the workspace. In Figure 1, Bayesian marginal posterior quantiles from (4) are compared to similar quantities for the frequentist third-order approximation to the p-value.

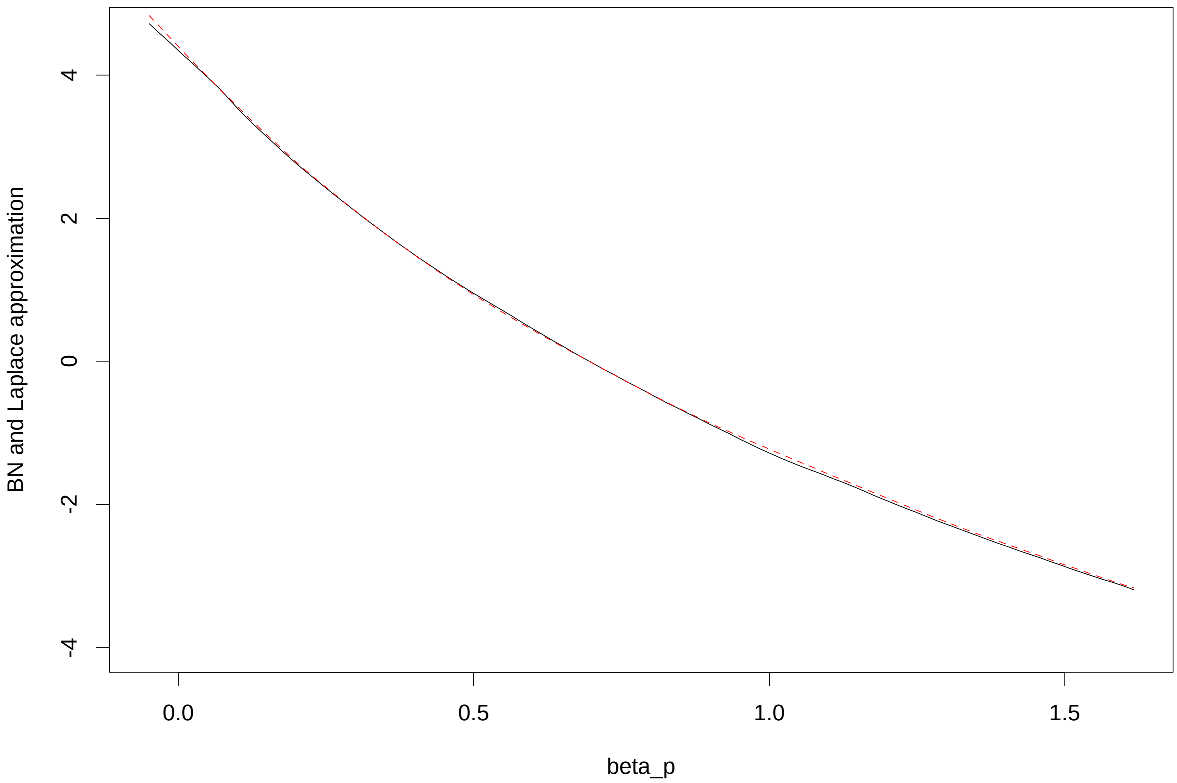

Figure 1: Modified likelihood root for third order frequentist approximation (solid line) and posterior quantile from approximation (4) as an approximation to the marginal posterior for β_6 (dotted line).

5.3 Random effects model

Consider the one-way random effects model $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$, for $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$, where τ_i and ϵ_{ij} are mutually independent with $\tau_i \sim N(0, \sigma_\tau^2)$ and

$\epsilon_{ij} \sim N(0, \sigma^2)$. For each i, the log-likelihood component is

$$\begin{aligned}
\ell(\mu, \sigma_\tau^2, \sigma^2; y_i) \;=\; & -\frac{1}{2}(n_i - 1)\log\sigma^2 - \frac{1}{2}\log(\sigma^2 + n_i\sigma_\tau^2) - \frac{1}{2}n_i\mu^2\ (\sigma^2 + n_i\sigma_\tau^2)^{-1} \\
& -\frac{1}{2}\frac{s_i^2}{\sigma^2} - \frac{1}{2}n_i\bar{y}_{i\cdot}^2\ (\sigma^2 + n_i\sigma_\tau^2)^{-1} + n_i\bar{y}_{i\cdot}\mu\ (\sigma^2 + n_i\sigma_\tau^2)^{-1}\ ,
\end{aligned}$$

where $\bar{y}_{i\cdot} = n^{-1}\sum_{j=1}^{n_i} y_{ij}$ and $s_i^2 = \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2$. Note this has the form of an exponential family log-likelihood, with some canonical parameters depending on the sample size. The log-likelihood function for the model is

$$\begin{aligned}
\ell(\mu, \sigma_\tau^2, \sigma^2) \;=\; & -\frac{1}{2}(N - k)\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{k}\log(\sigma^2 + n_i\sigma_\tau^2) - \frac{1}{2}\mu^2\sum_{i=1}^{k} n_i\ (\sigma^2 + n_i\sigma_\tau^2)^{-1} \\
& -\frac{1}{2}\sum_{i=1}^{k}\frac{s_i^2}{\sigma^2} - \frac{1}{2}\sum_{i=1}^{k} n_i\bar{y}_{i\cdot}^2\ (\sigma^2 + n_i\sigma_\tau^2)^{-1} + \mu\sum_{i=1}^{k} n_i\bar{y}_{i\cdot}\ (\sigma^2 + n_i\sigma_\tau^2)^{-1}\ ,
\end{aligned}$$

where $N = \sum_{i=1}^{k} n_i$.

If $\psi = \mu$ is the parameter of interest, $\eta = (\sigma_\tau^2, \sigma^2)^T$ is orthogonal to μ and a unique matching prior is obtained from (5). However the (ψ, ψ) component of the expected Fisher information matrix is a function only of the nuisance parameter:

$$i_{\psi\psi}(\psi, \sigma_\tau^2, \sigma^2) \propto \sum_{i=1}^{k} n_i\ (\sigma_\tau^2 + n_i\sigma^2)^{-1},$$

which, for $\eta_1, \eta_2 > 0$, meets the regularity assumptions stated in Lemma 1. Therefore we can further simplify the unique matching prior for $\psi = \mu$ to the flat prior:

$$\pi_U(\psi, \eta) \propto 1.$$

When $\psi = \sigma^2$ is the parameter of interest with $\eta = (\sigma_\tau^2, \mu)^T$ being the nuisance component, the orthogonality equations are more complicated. We can take $\lambda_2 = \sigma_\tau^2$;

note that $\hat{\lambda}_{2,\psi} = \bar{y}_{..}$. Then the differential equation (6) can be used to obtain λ_1.

Things simplify significantly for the balanced design $n_1 = \ldots = n_k = n$. The score functions corresponding to the nuisance parameter η have the form:

$$\begin{aligned} \ell_{\eta_1}(\psi, \eta) &= (\psi + n\eta_1)^{-2} \left\{ -\frac{nk}{2}\,(\psi + n\eta_1) + \frac{n^2}{2}\,\sum_{i=1}^{k}(\bar{y}_{i\cdot} - \eta_2)^2 \right\} \\ \ell_{\eta_2}(\psi, \eta) &= nk(\psi + n\eta_1)^{-1}\,\{\bar{y}_{\cdot\cdot} - \eta_2\}\,. \end{aligned}$$

By making use of Lemma 4, we identify $\lambda_1 = \psi + n\eta_1$ and $\lambda_2 = \eta_2$ as being orthogonal to the interest parameter ψ. Moreover for this reparameterization we have strong orthogonality: $\hat{\lambda}_{1,\psi} = \hat{\lambda}_1 = \frac{n}{k}\sum_{i=1}^{k}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$ and $\hat{\lambda}_{2,\psi} = \hat{\lambda}_2 = \bar{y}_{\cdot\cdot}$ for all ψ.

Another way to obtain the orthogonal reparameterization is by using the mean value transformation; when all n_i's are equal the random effects model is a full exponential family model, with canonical parameters $(\psi^{-1}, (\psi + n\eta_1)^{-1}, \eta_2(\psi + n\eta_1)^{-1})$.

Regardless of the method used, we find the partial information for ψ, $i_{\psi\psi.\eta}(\psi, \eta) \propto \psi^{-2}$, and the Jacobian of the transformation $|\partial\lambda/\partial\eta^T| = n$. Then by using (7) we obtain the prior:

$$\pi_U(\psi, \eta) \propto \psi^{-1} \tag{14}$$

which gives unique approximate matching inference based on matching priors in the orthogonal parameterization (ψ, λ^T).

The prior proposed by Levine & Casella (2003) gives the same approximate Bayesian inference as (14). However, the Bayesian inference associated with their prior is significantly more computationally intensive than the methods suggested here.

We performed a simulation study following Levine & Casella (2003). We randomly generated 100,000 data sets from the random effects model with $n = 10$ and $k = 3$, for $\mu = 10$ and $\sigma_\tau = \sigma = 1$. For each simulated data set, 95% posterior intervals were constructed for $\psi = \sigma^2$, by using the approximation to marginal posterior probability as given by (4). The computational work involved calculating this approximate tail probability for a grid of 200 values for ψ in $(0.01, 3)$. The posterior interval was easily obtained by spline smoothing. The simulated coverage of the 95% posterior intervals was 94.991%; the coverage obtained by Levine & Casella using a Metropolis Hastings algorithm with the prior $\pi(\psi, \eta) \propto \psi^{-1}(\psi + n\eta_1)^{-1}$ was 92.3%. The accuracy of our approach was also confirmed by the shape of the approximate marginal posterior probability function; the Bayesian p-value function for ψ, as approximated by (4), is almost identical to the exact one given by the chi-square distribution.

5.4 More theoretical examples

In this section we obtain first order matching priors for some examples discussed in Datta & Ghosh (1995) and Sweeting (2005).

Example 1. Inverse Gaussian model. Suppose that $Y_i \sim \mathrm{IG}(\mu, \sigma^2)$ with pdf

$$f(y; \mu, \sigma^2) \;=\; \frac{y^{-3/2}}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y-\mu)^2}{2\sigma^2\mu^2 y} \right\}, \qquad y > 0,$$

where $\mu > 0$ and $i = 1, \ldots, n$. This parameterization is orthogonal and the expected information matrix is $i(\mu, \sigma^2) = \mathrm{diag}(\mu^{-3}\sigma^{-2},\ \sigma^{-4}/2)$. When $\psi = \sigma^2$ and $\lambda = \mu$ we have strong orthogonality; $\hat{\lambda}_\psi = \hat{\lambda} = \bar{y}$, where $\bar{y} = n^{-1}\sum_{i=1}^n y_i$. Hence all the first order matching priors lead to unique approximation to the marginal posterior distribution as given by (2), and the unique matching prior is $\pi_U(\sigma^2, \mu) \propto \sigma^{-2}$. When the interest parameter is $\psi = \mu$ we do not have strong orthogonality any longer; $\hat{\lambda}_\psi = n^{-1}\sum_{i=1}^n y_i^{-1} + \bar{y}\psi^{-2} - 2\psi^{-1}$. The unique matching prior (5) is $\pi_U(\mu, \sigma^2) \propto \mu^{-3/2}\sigma^{-1}$. Datta & Ghosh (1995) propose the reverse reference prior $\pi_{RR}(\mu, \sigma^2) \propto \mu^{-3/2}\sigma^{-2}$, as it is a matching prior for each parameter in turn. This prior is of the form (1) with $g(\lambda) = \lambda^{-1/2}$, so both priors $\pi_{RR}(\mu, \sigma^2)$ and $\pi_U(\mu, \sigma^2)$ result in the same approximate Bayesian inference to order $O(n^{-1})$.

Example 2. Multivariate Normal mean. Suppose that $Y_i \sim N(\mu_i, 1)$ with $\mu_i \in \mathbb{R}$ for $i = 1, \ldots, p$, and take the parameter of interest to be $\psi = (\mu_1^2 + \ldots + \mu_p^2)^{1/2}$. Datta & Ghosh (1995) use the reparameterization $(\psi, \lambda_1, \ldots, \lambda_{p-1})$ with $\mu_1 = \psi \cos \lambda_1$, $\mu_2 = \psi \sin \lambda_1 \cos \lambda_2$, ... , $\mu_{p-1} = \psi \prod_{i=1}^{p-2} \sin \lambda_i \cos \lambda_{p-1}$, and $\mu_p = \psi \prod_{i=1}^{p-2} \sin \lambda_i \sin \lambda_{p-1}$; the information in this reparameterization is $i(\psi, \lambda) = \mathrm{diag}(1, \psi^2, \psi^2 \sin^2 \lambda_1, \ldots, \psi^2 \prod_{i=1}^{p-2} \sin^2 \lambda_i)$. This reparameterization also gives strong orthogonality, as we now show. The constrained maximum likelihood estimate $\hat{\lambda}_{p-1,\psi}$ is the solution of $\ell_{\lambda_{p-1}}(\psi, \lambda) = 0$, where

$$\ell_{\lambda_{p-1}}(\psi, \lambda) \propto y_{p-1} \sin \lambda_{p-1} - y_p \cos \lambda_{p-1},$$

yielding

$$\hat{\lambda}_{p-1,\psi} = \hat{\lambda} = \arctan \frac{y_p}{y_{p-1}}.$$

Next, we note that the score function corresponding to coordinate λ_{p-2}, $\ell_{\lambda_{p-2}}(\psi, \lambda)$ has the form

$$\ell_{\lambda_{p-2}}(\psi, \lambda) \propto y_{p-2} \sin \lambda_{p-2} - (y_{p-1} \cos \lambda_{p-1} + y_p \sin \lambda_{p-1}) \cos \lambda_{p-2},$$

and therefore the solution $\hat{\lambda}_{p-2,\psi}$ of the score equation $\ell_{\lambda_{p-2}}(\psi, \lambda) = 0$ is

$$\hat{\lambda}_{p-2,\psi} = \hat{\lambda}_{p-2} = \arctan \frac{y_{p-1}}{y_{p-2}} \cos \hat{\lambda}_{p-1} + \frac{y_p}{y_{p-2}} \sin \hat{\lambda}_{p-1}.$$

We continue with backward procedure, which resembles the one introduced in the

Appendix and obtain $\hat{\lambda}_\psi = \hat{\lambda}$. Having strong orthogonality, the unique matching prior is $\pi_U(\psi, \lambda) \propto 1$. Datta & Ghosh (1995) and Tibshirani (1989) obtained $\pi_R(\psi, \lambda) \propto \Pi_{k=1}^{p-1} \sin^{p-1-k} \lambda_k$ as a first order matching prior for ψ; this prior is also a reference prior. Both priors give the same posterior quantiles to third order.

Example 3. Normal mean product. Suppose that we have $Y_i \sim N(\mu_i, 1)$ independent variables with $i = 1, 2$. As in Datta & Ghosh (1995) we consider the orthogonal parameterization $\psi = 2\mu_1\mu_2$, $\lambda = \mu_1^2 - \mu_2^2$. The unique matching prior is $\pi_U(\psi, \eta) \propto (\psi^2 + \lambda^2)^{-1/4}$, which is also the reference prior obtained by Datta & Ghosh.

Example 4. Exponential mean ratio. Let Y_1 and Y_2 be independent exponential random variables with means η and $\psi\eta$ respectively. A version of the orthogonal parameter is $\lambda = \psi\eta^2$ and the information matrix is $i(\psi, \lambda) = \mathrm{diag}(\psi^{-2},\ \frac{1}{2}\lambda^{-2})$. Our approach suggests the unique matching prior $\pi_U(\psi, \lambda) \propto \psi^{-1}$; in terms of the original parameterization $\pi_U(\psi, \eta) \propto \eta$. The reference prior $\pi_R(\psi, \lambda) \propto \psi^{-1}\lambda^{-1}$ is a matching prior not only for ψ but also for λ and leads to the same approximate Bayesian inference to order $O(n^{-1})$.

Example 5. Normal coefficient of variation. Suppose that $Y_i \sim N(\mu, \sigma^2)$, for $i = 1, \dots, n$ and assume $\psi = \sigma^{-1}\mu$ is the parameter of interest, while $\eta = \sigma$. The information matrix for this example is given in Sweeting (2005) as

$$i(\psi, \eta) = \left(\begin{array}{cc} 1 & \eta^{-1}\psi \\ \eta^{-1}\psi & 2\eta^{-2}(1 + \frac{1}{2}\psi) \end{array} \right),$$

giving the partial information for ψ as $i_{\psi\psi.\eta}(\psi, \eta) \propto (1 + \frac{1}{2}\psi^2)^{-1}$. An orthogonal component for ψ is $\lambda = \eta\ (1 + \frac{1}{2}\psi^2)^{1/2}$ and the unique matching prior in the initial parameterization, obtained by (7), is $\pi_U(\psi, \eta) \propto 1$.

Example 6. Log-normal mean. Suppose that $Y_i \sim N(\mu, \sigma^2)$ and the interest parameter is $\psi = \mu + \frac{1}{2}\sigma^2$, with nuisance parameter $\eta = \sigma^2$. We use the information matrix given in Sweeting (2005) to calculate the partial information for ψ as $i_{\psi\psi.\eta}(\psi, \eta) = \eta^{-1}(1 + \frac{1}{2}\eta)^{-1}$. The orthogonal component is $\lambda = \psi - \frac{1}{2}\eta - \log \eta$; Datta & Ghosh (1995) describe how to construct the orthogonal parameterization for this model. The unique matching prior, in the original parameterization, is given by $\pi_U(\psi, \eta) \propto \eta^{-3/2}(1 + \frac{1}{2}\eta)^{1/2}$.

Example 7. Weibull distribution. Consider a sample $Y_1, \ldots, Y_n$ distributed according to the density function:

$$f(y; \phi) = \beta\eta\ (\eta y)^{\beta-1}\ \exp\{-(\eta y)^{\beta}\} \qquad \text{where} \quad y > 0 \text{ and } \beta, \eta > 0.$$

For the interest parameter $\psi = \beta$ an orthogonal component, as derived in Severini (2000, Ch 4.5), is $\lambda = \log \eta + (\gamma - 1)/\psi$, where γ denotes Euler's constant. The information matrix in terms of the orthogonal reparameterization is $i(\psi, \lambda) \propto \text{diag}\,(\psi^{-2}(\pi^2/6 - 1),\ \psi^2)$. Thus the unique prior, expressed in the original parameterization is $\pi_U(\psi, \eta) \propto \psi^{-1}\eta^{-1}$.

6 Acknowledgement

The authors would like to acknowledge helpful discussions with Anthony Davison, Don Fraser and Rahul Mukerjee. This research was partially supported by Natural Sciences and Engineering Research Council of Canada.

A Appendix

A.1 Invariance of q_B

Lemma 1 *Assume that $g(\lambda)$ is a smooth function such that $g(\lambda) \neq 0$ and $g'(\lambda)$ is continuous for all λ. If we denote by q_B^g the q_B quantity from (3) under the matching prior $\pi^g(\psi, \lambda) \propto i_{\psi\psi}^{1/2}(\psi, \lambda)\; g(\lambda)$, then q_B^g can be approximated with second order relative error, without detailed specification of the function $g(\lambda)$.*

Proof. For simplicity, we prove the lemma for scalar λ; the vector λ case is shown in a similar way. Due to the parameter orthogonality when ψ is in its moderate deviation region $\psi = \hat{\psi} + n^{-1/2}\Delta_\psi$, we have $\hat{\lambda}_\psi = \hat{\lambda} + n^{-1}\Delta_{\psi,\lambda}$, where $\Delta_{\psi,\lambda} = O_p(1)$. More precisely, if we denote by $\theta_0 = (\psi_0, \lambda_0)$ the true parameter value, Cox & Reid (1989) show that for fixed Δ_ψ we have:

$$\Delta_{\psi,\lambda} = -\frac{n^{1/2} j_{\psi\lambda}(\theta_0)}{i_{\lambda\lambda}(\theta_0)}\Delta_\psi + \frac{1}{2}\frac{i_{\psi^2\lambda}(\theta_0)}{i_{\lambda\lambda}(\theta_0)}\Delta_\psi^2 + O_p(n^{-1/2}),$$

where $i_{\psi^2\lambda}(\theta_0) = n^{-1}E[\partial^3\ell(\theta)/\partial\psi^2\partial\lambda; \theta_0]$.

Applying the mean value theorem to $g(\lambda)$ we find

$$g(\hat{\lambda}_\psi) = g(\hat{\lambda}) + n^{-1}\; \Delta_{\psi,\lambda}\; g'(\xi_{\psi,\lambda})$$

where $\xi_{\psi,\lambda}$ is between $\hat{\lambda}$ and $\hat{\lambda}_\psi$; as $g(\lambda) \neq 0$ for all λ we can also write:

$$\frac{g(\hat{\lambda}_\psi)}{g(\hat{\lambda})} = 1 + n^{-1}\Delta_{\psi,\lambda}\frac{g'(\xi_{\psi,\lambda})}{g(\hat{\lambda})}. \tag{A-1}$$

Since $g'(\lambda)$ is continuous, for $\|\xi_{\psi,\lambda} - \hat{\lambda}\| < n^{-1}|\Delta_{\psi,\lambda}|$ with $\Delta_{\psi,\lambda} = O_p(1)$ we obtain that $g'(\xi_{\psi,\lambda}) - g'(\hat{\lambda}) \to 0$ in probability as $n \to \infty$. Dividing by $g(\hat{\lambda}) \neq 0$ gives $g'(\xi_{\psi,\lambda})/g(\hat{\lambda}) - g'(\hat{\lambda})/g(\hat{\lambda}) \to 0$ in probability as $n \to \infty$. We consider models for which the regularity assumptions, that require consistency of the maximum likelihood estimators, are met. In other words, under the regularity assumptions we have that $\hat{\lambda} \to \lambda_0$ in probability as $n \to \infty$. Now, using the smoothness of function g that assures the ratio g'/g to be continuous we ascertain that $g'(\xi_{\psi,\lambda})/g(\hat{\lambda}) \to g'(\lambda_0)/g(\lambda_0)$ in probability as $n \to \infty$, which implies:

$$\frac{g'(\xi_{\psi,\lambda})}{g(\hat{\lambda})} = O_p(1).$$

Therefore we conclude that in (A-1) the ratio $g(\hat{\lambda}_\psi)/g(\hat{\lambda})$ can be approximated to $O_p(n^{-1})$ accuracy, and thus q_B is invariant to error of order $O_p(n^{-1})$ under Peers-Tibshirani priors. In particular, if $\tilde{q}_B$ equals q_B under the prior $\pi_U(\psi, \lambda) \propto i_{\psi\psi}^{1/2}(\psi, \lambda)$, we have:

$$q_B^g = \tilde{q}_B\{1 + O_p(n^{-1})\}.$$

The same reasoning shows the invariance to $O_p(n^{-1})$ of the marginal posterior density under Peers-Tibshirani priors. The approximation to q_B^g enables further a similar approximation to the marginal posterior tail probability $1 - \Pi_m^g(\psi \mid y)$.

A.2 Invariance of $\Pi_m^g(\psi \mid y)$ under matching priors

Lemma 2 *Under the above assumptions for $g(\lambda)$, all the first order matching priors $\pi^g(\psi, \lambda)$ of the form (1) result in marginal posterior probabilities $\Pi_m^g(\psi \mid y)$ that have a unique approximation to $O(n^{-1})$, when $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$.*

Proof. To show this result we use approximation (4), to the marginal posterior $\Pi_m^g(\psi \mid y)$, and write $r^* = r - r^{-1}\log(r/q_B)$. Denote by r^g and r^{*g} the signed likelihood root and r^* statistic respectively, corresponding to prior $\pi^g(\theta)$, and by $\tilde{r}$ and $\tilde{r}^*$ the same quantities corresponding to prior $\pi_U(\theta)$.

Using Lemma 1, when $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$ we express q_B^g as $q_B^g = \tilde{q}_B\{1 + n^{-1}C\}$, where $C = O_p(1)$. We show first that a similar relationship between r_B^{*g} and $\tilde{r}_B^*$ holds. For this we consider the asymptotic expansion of the signed likelihood root r in terms of q_B (see Reid, 2003) $r = q_B + n^{-1/2}Aq_B^2 + n^{-1}Bq_B^3$, with $A, B = O_p(1)$ which gives further:

$$r^* = q_B + \frac{A}{n^{1/2}}(q_B^2 + 1) + \frac{1}{n}\left\{Bq_B^3 + (B - 3A^2/2)q_B\right\}.$$

Writing such expansions for both r_B^{*g} and $\tilde{r}^*$ and using (A-2) we find that $r^{*g} = \tilde{r}^* + n^{-1}C\tilde{q}^B + O_p(n^{-3/2})$, or:

$$r^{*g} = \tilde{r}^*(1 + n^{-1}C), \tag{A-2}$$

to order $O_p(n^{-3/2})$. Next we prove a similar result for the standard normal approximation of r^{*g} and $\tilde{r}^*$ respectively.

The Taylor series expansion of $\Phi(x(1 + n^{-1}c)$ with respect to x gives

$$\begin{aligned} \Phi\left\{x(1 + \frac{c}{n})\right\} &= \Phi(x) + \frac{c}{n}x\ \phi(x) \\ &= \Phi(x)\{1 + O(n^{-1})\} \qquad \text{for all } x > 0 \end{aligned} \tag{A-3}$$

to $O(n^{-3/2})$. The relative error is a consequence of the ratio $x\phi(x)/\Phi(x)$ being bounded for all non-negative values; the range of the ratio for all $x \geq 0$ is given by the range corresponding to $0 \leq x \leq 1$.

Using (A-3) with (A-2) gives

$$\Phi(r^{*g}) = \Phi(\tilde{r}^*)\ (1 + n^{-1}D) \qquad \tilde{r}^* \geq 0, \tag{A-4}$$

where $D = O_p(1)$. For $\tilde{r}^* < 0$, one needs to calculate the tail probability instead of cumulative distribution at value $\tilde{r}^*$. More precisely same result can be established for $1 - \Phi(\tilde{r}^*)$, by using (A-4) for $-\tilde{r}^*$ and then the fact $\Phi(-\tilde{r}^*) = 1 - \Phi(\tilde{r}^*)$.

We conclude that for smooth functions $g(\lambda)$ all matching priors of form (1) lead to unique approximation to the marginal posterior tail probability $1 - \Pi_m(\psi \mid y)$, with relative error of at most $O(n^{-1})$, as given by $\Phi(\tilde{r}^*)$, or its asymptotically equivalent version (2).

The result was proved in DiCiccio & Martin (1993) by using the relationship between q_B (denoted by T in their paper) and the variable U introduced by Barndorff-Nielsen (1986) which, when used in an expression of form (2) or (4), provides inference accurate to order $O(n^{-3/2})$. They establish that in the orthogonal parameterization we have $q_B = U + O_p(n^{-1})$, which added to results of Barndorff-Nielsen (1986) demonstrate the second order accuracy of p-values and confidence limits obtained by using q_B approach with any matching prior. The exact invariance of q_B under reparameterization only, in addition to the approximate invariance to error of order $O_p(n^{-1})$ under the choice of matching priors, extends the result to models in a general parameterization.

A.3 Invariance of posterior quantiles $\hat{\psi}^{(1-\alpha)}(\pi^g, y)$ under matching priors

Lemma 3 *Assume that $g(\lambda)$ is a smooth function as before. Let $0 < \alpha < 1$ and denote by $\hat{\psi}^{(1-\alpha)}(\pi^g, y)$ the posterior quantile corresponding to prior $\pi^g(\theta) \propto i^{1/2}_{\psi\psi}(\theta)\, g(\lambda)$, which is defined by $\Pi^g_m\{\hat{\psi}^{(1-\alpha)}(\pi^g, y) \mid y\} = 1 - \alpha$. Then*

$$\hat{\psi}^{(1-\alpha)}(\pi^g, y) = \hat{\psi}^{(1-\alpha)}(\pi_U, y) + O_p(n^{-3/2}); \qquad \text{(A-5)}$$

that is, the posterior quantile is unique to $O_p(n^{-3/2})$ *under the class of matching priors* $\pi^g(\theta)$.

Proof. Let z_α denote the $100(1-\alpha)$ percentile point of a standard normal variate and let $j^{\psi\psi}(\psi,\lambda)$ stand for the (ψ,ψ) component of the inverse of the observed information matrix. Cornish-Fisher inversion of the Edgeworth expansion for the marginal posterior distribution of ψ leads to:

$$\begin{aligned}\hat{\psi}^{(1-\alpha)}(\pi^g,y) &= \hat{\psi}+n^{-1/2}\{j^{\psi\psi}(\hat{\theta})\}^{1/2}\ z_\alpha \\ &\quad + n^{-1}\{j^{\psi\psi}(\hat{\theta})\}^{1/2}\ u_1(z_\alpha,\pi^g,y)+O_p(n^{-3/2}),\end{aligned}$$

where $u_1(z_\alpha,\pi^g,y)=A_{11}(\pi^g,y)+A_{12}(y)+(z_\alpha^2+2)A_3(y)$ with

$$A_{11}(\pi^g,y) = \{j^{\psi\psi}(\hat{\theta})\}^{-1/2}\left\{\frac{\pi^g_\psi(\hat{\theta})}{\pi^g(\hat{\theta})}\ j^{\psi\psi}(\hat{\theta})+\frac{\pi^g_{\lambda^T}(\hat{\theta})}{\pi^g(\hat{\theta})}\ j^{\lambda\psi}(\hat{\theta})\right\} \tag{A-6}$$

$\pi^g_\psi(\theta)=\partial\pi^g(\theta)/\partial\psi$, $\pi^g_\lambda(\theta)=\partial\pi^g(\theta)/\partial\lambda$ and expressions for A_{12} and A_3 are given in Mukerjee & Reid (1999). To prove that $\hat{\psi}^{(1-\alpha)}(\pi^g,y)$ is unique to $O_p(n^{-3/2})$ under the class of matching priors $\pi^g(\theta)$, it amounts to show that $u_1(z_\alpha,\pi^g,y)$, or furthermore $A_{11}(\pi^g,y)$ does not depend on $g(\lambda)$ to order $O_p(n^{-1/2})$.

Taking the partial derivatives of the prior $\pi^g(\theta)$

$$\begin{aligned}\pi^g_\psi(\theta) &= g(\lambda)\frac{\partial}{\partial\psi}\{i_{\psi\psi}(\theta)\}^{1/2} \\ \pi^g_\lambda(\theta) &= g(\lambda)\frac{\partial}{\partial\lambda}\{i_{\psi\psi}(\theta)\}^{1/2}+\{i_{\psi\psi}(\theta)\}^{1/2}\frac{\partial}{\partial\lambda}g(\lambda)\end{aligned}$$

we remark that the factor in braces in (A-6) simplifies to:

$$
\begin{aligned}
&\{i_{\psi\psi}(\hat{\theta})\}^{-1/2}\left[\frac{\partial}{\partial\psi}\{i_{\psi\psi}(\hat{\theta})\}^{1/2}\right] j^{\psi\psi}(\hat{\theta}) \\
&+ \{i_{\psi\psi}(\hat{\theta})\}^{-1/2}\left[\frac{\partial}{\partial\lambda^T}\{i_{\psi\psi}(\hat{\theta})\}^{1/2}\right] j^{\lambda\psi}(\hat{\theta}) + \{g(\hat{\lambda})\}^{-1}\left\{\frac{\partial}{\partial\lambda^T} g(\hat{\lambda})\right\} j^{\lambda\psi}(\hat{\theta}).
\end{aligned}
$$

This remark concludes our proof that (A-5) holds, as the sum of the first two terms corresponds to

$$
\frac{\partial/\partial\psi\{\pi_U(\hat{\theta})\}}{\pi_U(\hat{\theta})}\, j^{\psi\psi}(\hat{\theta}) + \frac{\partial/\partial\lambda^T\{\pi_U(\hat{\theta})\}}{\pi_U(\hat{\theta})}\, j^{\lambda\psi}(\hat{\theta}),
$$

while the last term is $O_p(n^{-1/2})$ since $j^{\lambda\psi}(\hat{\theta}) = O_p(n^{-1/2})$, because of parameter orthogonality, and $g_{\lambda^T}(\hat{\lambda})/g(\hat{\lambda}) = O_p(1)$ due to the assumptions for $g(\lambda)$; see also Lemma 1.

A.4 On strong orthogonality: vector nuisance parameter

Lemma 4 *Assume the score function for the parametric model $f(y;\psi,\eta)$ has the form*

$$
\begin{aligned}
\ell_{\eta_1}(\psi,\eta;\ y) &\propto h_1\{\lambda_1(\psi,\eta_1);\ y\} && \text{(A-7)} \\
\ell_{\eta_2}(\psi,\eta;\ y) &\propto h_2\{\lambda_2(\psi,\eta_1,\eta_2),\ h_1(\cdot)f_{2,1}(\cdot);\ y\} && \text{(A-8)}
\end{aligned}
$$

and, for a general $k = 2, 3, \ldots, d-1$, of the form:

$$
\ell_{\eta_k}(\phi;\ y) \propto h_k\{\lambda_k(\psi,\eta_1,\ldots\eta_k),\ h_1(\cdot)f_{k,1}(\cdot),\ldots,\ h_{k-1}(\cdot)f_{k,k-1}(\cdot);\ y\} \qquad \text{(A-9)}
$$

where the proportionality refers to factors that are non trivial functions of parameters only. Then the reparameterization (ψ,λ^T) is an orthogonal one, with $\hat{\lambda}_\psi = \hat{\lambda}$.

Proof. We first note that in all expressions (A-7) to (A-9) the dependence of the components h_k is via products of terms involving only $h_1, \ldots, h_{k-1}$. At the same time, more freedom is allowed for the dependence of h_k on other quantities $f_{k,l}(\cdot)$, although it is only through products of form $h_l(\cdot)f_{k,l}(\cdot)$ with $l < k$. Similarly, the functions h_k may depend on $\lambda_1, \ldots, \lambda_{k-1}$ alone for $k = 2, \ldots, d-1$.

To show that for these models, in the (ψ, λ^T) parameterization, we have $\hat{\lambda}_\psi = \hat{\lambda}$, we need to calculate the constrained maximum likelihood estimates $\hat{\lambda}_\psi$ for each component in part. The particular dependence of the score function for η on the components η_k yields $\lambda_1 = \lambda_1(\psi, \eta_1)$, $\lambda_k = \lambda_k(\psi, \eta_1, \ldots, \eta_k)$ for $k = 2, \ldots, d-1$, thus underlining a forward procedure to determine them. The approach is illustrated by Example 2 of §5.4.

We begin by solving the score equation corresponding to the first component of $\hat{\eta}_\psi$, $\hat{\eta}_{1,\psi}$. Using (A-7) the equation $\ell_{\eta_1}(\psi, \hat{\eta}_\psi) = 0$ is equivalent to

$$h_1(\lambda_1;\ y) = 0 \qquad \text{for } \hat{\lambda}_{1,\psi} = \lambda_1(\psi, \hat{\eta}_{1,\psi}).$$

The solution is $\hat{\lambda}_{1,\psi}$ is free of ψ because it depends only on y; hence $\hat{\lambda}_{1,\psi} = \hat{\lambda}_1$. We continue next with the score equation for the subsequent nuisance parameter η_2, and equate it to 0, i.e. $\ell_{\eta_2}(\psi, \hat{\eta}_\psi) = 0$. From (A-8) this is equivalent to solving

$$h_2\left\{\lambda_2,\ h_1(\hat{\lambda}_{1\psi}; y) f_{2,1}(\psi, \hat{\eta}_\psi);\ y\right\} = 0 \qquad \text{for } \hat{\lambda}_{2,\psi} = \lambda_2(\psi, \hat{\eta}_{1,\psi}, \hat{\eta}_{2,\psi}).$$

However, since $h_1\left\{\lambda_1(\psi, \hat{\eta}_{1,\psi})\right\} = 0$, the equation simplifies to $h_2(\lambda_2,\ 0) = 0$, which of course has a solution $\hat{\lambda}_{2,\psi} = \lambda_2(\psi, \hat{\eta}_{1,\psi}, \hat{\eta}_{2,\psi})$ free of ψ. By mathematical induction one deduces that $\hat{\lambda}_{k,\psi} = \hat{\lambda}_k$ for each $k = 1, \ldots, d-1$.

Lemma 4 states that for all the models $f(y; \psi, \eta)$ for which the score equations corresponding to the nuisance parameter η can be written in the form (11), or more generally (A-9), the orthogonal component does not have to be constructed from the partial differential equations. The orthogonal component is revealed from the score equation for the nuisance parameter.

References

[1] Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307-22.

[2] Barndorff-Nielsen, O.E. & Cox, D.R. (1994). *Inference and Asymptotics.* London: Chapman & Hall/CRC.

[3] Brazzale, A.R. (2000). Practical small-sample parametric inference. Ph.D. Thesis N. 2230, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne.

[4] Cox, D.R. & Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with Discussion). *J. R. Statist. Soc.* B **49**, 1-39.

[5] Cox, D.R. & Reid, N. (1989). On the stability of the maximum-likelihood estimators of orthgonal parameters. *Can. J. Statist.* **17**, 229-33.

[6] Davison, A.C. & Hinkley, D.V. (1997). *Bootstrap Methods and Their Application.* Cambridge: Cambridge University Press.

[7] Datta, G.S. & Ghosh, M. (1995). Some remarks on noninformative priors. *J. Am. Statist. Assoc.* **90**, 1357-63.

[8] Datta, G.S. and Mukerjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics. Lecture Notes in Statistics* **178**, New York: Springer-Verlag.

[9] DiCiccio, M.J. & Martin, M.A. (1991). Approximations of marginal tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference. *Biometrika* **78**, 891-902.

[10] DiCiccio, M.J. & Martin, M.A. (1993). Simple modifications for signed roots of likelihood ratio statistics. *J. R. Statist. Soc.* B **55**, 305-16.

[11] Levine, R.A. & Casella, G. (2003). Implementing matching priors for frequentist inference. *Biometrika* **90**, 127-37.

[12] Mukerjee, R. & Dey, D.K. (1993). Frequentist validity of posterior quantiles in the presence of a nuisance parameter: higher order asymptotics. *Biometrika* **80**, 499-505.

[13] Mukerjee, R. & Ghosh, M. (1997). Second-order probability matching priors. *Biometrika* **84**, 970-75.

[14] Mukerjee, R. & Reid, N. (1999). On a property of probability matching priors: Matching the alternative coverage probabilities. *Biometrika* **86**, 333-40.

[15] Nicolaou, A.(1993). Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. *J. R. Statist. Soc.* B **55**, 377-90.

[16] Peers, H.W. (1965). On confidence points and Bayesian probability points in the case of several parameters. *J. R. Statist. Soc. B* **27**, 9-16.

[17] Reid, N. (2003). Asymptotics and the theory of inference. *Annals of Statistics* **31**, 1695-731.

[18] Reid, N., Mukerjee, R. & Fraser, D.A.S. (2003). Some aspects of matching priors. In *Mathematical Statistics and Applications: Festschrift for Constance van Eeden*, Eds. M. Moore, S. Froda and C. Leger, pp 31-44, CRM, Montreal.

[19] Severini, T.A. (2000). *Likelihood Methods in Statistics.* Oxford: Oxford University Press.

[20] Sweeting, T.J. (2005). On the implementation of local probability priors for interest parameters. *Biometrika* **92**, 47-57.

[21] Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika* **76**, 604-8.

[22] Tierney, L & Kadane J.B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Assoc.* **81**, 82-6.

[23] Welch, B.L. & Peers, H.W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. R. Statist. Soc.* B **25**, 318-29.